EXHIBIT B

Partial Translation of "A Speaker Verification Method Which can Control False Acceptance Rate" in "THE TRANSACTIONS OF THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS D-II" by S. HAYAKAWA, K. TAKEDA, and F. ITAKURA, published in December, 1999

(Page 2213, left column, line 5 – page 2214, right column, line 2)

2. Speaker verification method by distribution of interspeaker distances

The speech of a person concerned is varied on the utterance basis; however, the relative relationship between the person concerned and another speaker is considered to be stable, compared with the fluctuation in the speaker. Matsui et al. performed a speaker verification experiment of a text independent type, using speech data actually having a difference in utterance period, and showed that an FR rate is more largely influenced by the fluctuation in utterance due to the difference in period, compared with an FA rate. This result suggests that, in the speaker verification, the determination of "the possibility that a speaker is not others" with a certain risk is more stable with respect to the fluctuation on the utterance basis, compared with the determination of "the possibility that a speaker is the one concerned".

A method for verifying a speaker is proposed. According to this method, in speaker verification, a distance with respect to another speaker, as well as a distance with respect to a claimed speaker are obtained, and the probability distribution thereof is estimated to obtain a probability at which the claimed speaker is included in a group of other speakers, whereby a speaker is verified.

FIG. 1 shows a block diagram showing a proposed verification method. When speech data is input together with the claim that "this is a speaker K", the distance calculation is performed between the speech data and registered speeches of all the registered N speakers, whereby a distance d_i (i = 1, 2, ..., N) between the speaker k and each of the N speakers is

obtained. Next, a distribution $F(d; \theta)$ (hereinafter, referred to as a distribution of an interspeaker distance) of distances between a group of other speakers and the input speech is estimated from a collection $\{d_i \mid i \neq k\}$ of $(N-1)$ distances excluding a distance d_k with respect to the claimed speaker. Herein, θ is a distribution parameter. The final determination of acceptance/rejection is performed by comparing a probability value $F(d_k;\theta)$ at which the distance d_k with respect to the claimed speaker is output from the distribution $F(d;\theta)$ with a previously set FA rate.

The distribution of the interspeaker distances can be approximated with a complicated distribution, if a mixed gauss distribution is used. However, in this thesis, the distribution of the interspeaker distances is approximated with a single normal distribution that is easy to handle. FIG. 2 shows the distribution of the interspeaker distances in word data uttered by male speakers used in the experiment of Section 3. The distances are normalized with an average value and a standard deviation of the distances with respect to all the registered speakers, on the pretender's input basis. A solid line represents a probability density function of the normal distribution. It is understood from the figure that the outline of the distribution of interspeaker distances is substantially in accordance with the normal distribution.

In the case where the distribution of the interspeaker distances is approximated with the normal distribution, an average value μ and a standard deviation σ that are probability distribution parameters are obtained by the following expressions.

$$\mu = \frac{1}{N-1}\sum_{\substack{n=1 \\ n\neq k}}^{N} dn \qquad (1)$$

$$\sigma^2 = \frac{1}{N-2}\sum_{\substack{n=1 \\ n\neq k}}^{N} (d_n - \mu)^2 \qquad (2)$$

It should be noted that the claimed speaker is set to be k. The following verification expressions are configured using the average value μ

and the standard deviation σ of the interspeaker distances.

if $d_k < \mu - \alpha \cdot \sigma$, then accept,

if $d_k \geq \mu - \alpha \cdot \sigma$, then reject. (3)

where α is a normalized distance obtained by a normal probability distribution function:

$$\Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{\frac{-t^2}{2}dt} \tag{4}$$

corresponding to the previously specified FA rate. If the input speech is that of the claimed speaker, its distance (in the speaker) is considered to be necessarily a value smaller than an average value of the distribution of the interspeaker distances. Therefore, the distance is determined with a probability at which the distance is included only on one side of the normal distribution function. For example, in the case of setting the FA rate to be 5%, $1 - \Phi(\alpha) = 0.05$, i.e., $\alpha = 1.65$ is used.

Furui proposes a procedure for determining a threshold value of verification, using the relative stability of the distribution of the interspeaker distances, compared with the distribution of the distances in the speaker. According to this method, when the template of the speaker k is updated, based on the average value and variance of distances between the utterance data (which is held by a system) of (N−1) speakers excluding k and the template of the updated speaker k:

$$\mu k = \frac{1}{N-1} \sum_{\substack{n-1 \\ n \neq k}}^{N} d(k,n) \tag{5}$$

$$\sigma_k^2 = \frac{1}{N-2} \sum_{\substack{n=1 \\ n \neq k}}^{N} (d(k,n) - \mu k)^2 \tag{6}$$

the threshold value θ_k of the speaker k is updated by

$$\theta_k = \alpha(\mu_k - \sigma_k) + b \tag{7}$$

where d(k,n) represents the distance between the template of the speaker k and the utterance data of the speaker n, and a and b are constant parameters which are set to be common to all the speakers by a preliminary

experiment.

In the present thesis, the proposed method is similar to the procedure proposed by Furui, in that the distribution of interspeakers is used for determination, and they have the following two features:

(1) according to the present procedure, the interspeaker distances are calculated with the input speech of the claimed speaker to be verified, and verification is performed with a relative value thereof, whereby determination robust to the fluctuation due to the difference in period and the utterance environment can be expected in the same way as in the case of using a cohort model [6]; and

(2) without obtaining parameters such as a and b as in the procedure proposed by Furui, determination with a false acceptance rate (FA rate) being set is possible.

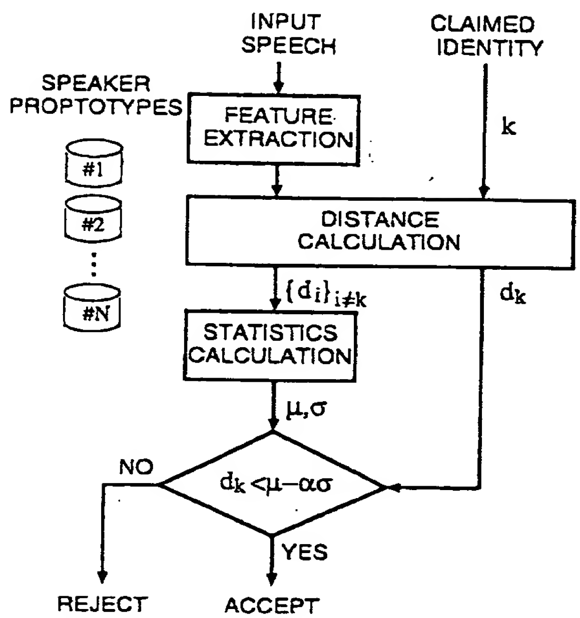

Fig. 1 Block diagram for proposed speaker verification method.

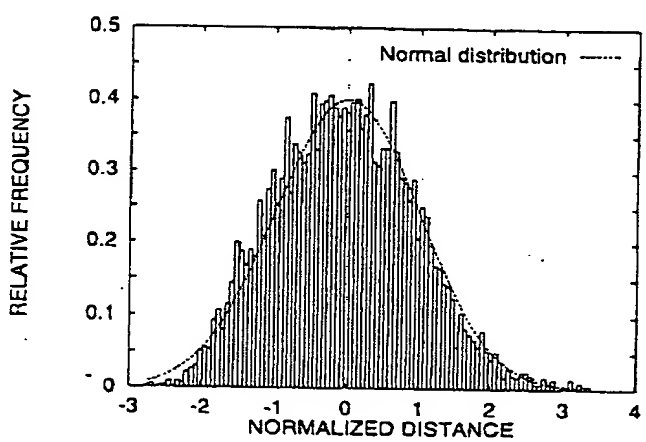

Fig. 2 Distribution of the interspeaker distances.